

Transcription Pattern for Mobile Call

James K. Tamgno, Loïc Akoha, Kibanou Eric Tidye

DTE-LTI, ESMT-UCAD, Dakar Senegal

james.tamgno@esmt.sn, akoick2001@hotmail.com, kibanou.tidiye@hotmail.fr

Abstract — These paper summarizes the research carried out as part of the design of an Android application to transcript phone calls. The system will automatically collect out any mobile calls and transcript it with voice recognition into text, which will therefor be reordered in the memory of each mobile phone of the callers. The system allows someone to make calls and send text messages generated to both. In order to archive this, we first of all analyze voice recognition models available on different platform (iOS and Android, BlackBerry, Nokia), then we design using Android features to build new patterns necessities to setup owner solution.

Keywords : *Android, Speech, GSM, Mobile System.*

I. INTRODUCTION

A- Context and motivations

1- Penetration of mobile telephony

The speech recognition systems have made significant progress in recent years. These advances have been made through research on acoustic and language models but also through the continued growth of processing power. This increase in power has not only get the maximum benefit from the algorithms used, but also can perform the speech recognition software on current PC configurations, which opened a wide market and, therefore, led to the development research.

Today, functional systems are based on a statistical approach. Market leaders typically offer software for recognition of continuous speech with the possibility of controls for common functions of word processing applications [1]. They advertise vocabulary sizes ranging from 30 000 to 60 000 words for language, allowing the driven gear to the 120 to 160 words per minute (which corresponds to a relatively high speed of words) with a higher success of recognition 95%. We will see later, during testing, it is in reality.

2- Millenium Development goals

The natural language processing is a vast field of research which combines specialists from many disciplines: linguists, computer scientists, logicians, psychologists, translators ... So it is an area economically holder whose numerous applications in areas as diverse as office, assisting the disabled, education, home automation, translation, navigation aid, documentation, not to mention computing. The use of speech recognition thus has a very wide range of application that requires mostly still research.

B- Circumstances

1- The situation

Basically, the voice recognition software used primarily to enter text in large quantities while passing the keyboard, providing an average rate of 50 words per minute for manual transcription, against more than 150, for speech. That said, it was a desktop-oriented and / or machinery that made necessary the use of the keyboard for text corrections on a computer solution.

The voice recognition applications are now declined on different platform (iOS and Android, BlackBerry, Nokia, etc.), And respond to different practical needs related to mobile use, especially in working life such as using hands-configuration-free for standard voice commands (search the directory, calls), translation, web searches.

It is in the context of our study to build on what has already been done to propose a new transcription service based mainly on the processing of telephone calls, all in an independent and autonomous unit. Indeed, like sms or photos, phone conversations leave little trace. Therefore, it may be useful to keep written records and / or voice phone calls, depending on the circumstances [2]. And some may be used as a proof of harassment, others to keep memories of birthday greetings, or perform transcription live telephone interviews for journalists to mention only those examples.

This application that we will develop will therefore propose a new solution to the problems of any kind encountered by different users, allowing them to adapt to their various needs [3].

2- Needs and expectations

Mobile applications have become over the years the elements of everyday life. Indeed, they offer the general public the continuity of services and tasks performed, based obviously on the mobility offered by smartphones, tablets and others. Therefore, the aim here is to provide a portable and ergonomic solution to allow different users to take advantage of speech recognition (Fig 1.).

To do this, our work will first explain in a clear and entertaining way the principles governing the operation of the speech recognition, and then to implement it in an Android application in order to achieve voice transcription service telephone conversations used to various purposes. In addition, this project should enable us to confront the workplace with different constraints.

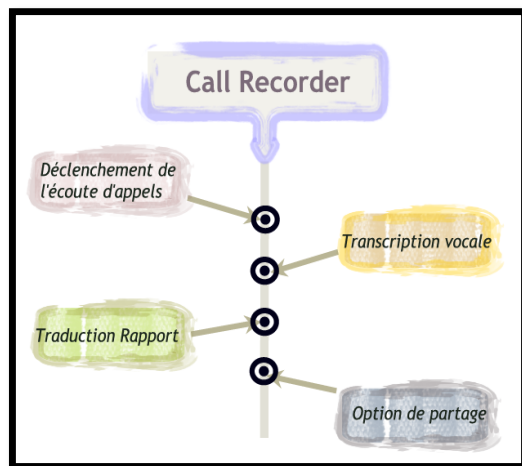


Fig 1. Application structure [4]

II. ANALYTICAL APPROACH

What is Speech Recognition?

Automatic speech recognition is an area of science that has always had great appeal among researchers as to the general public. Indeed, who has never dreamed of being able to talk with a machine or at least control a device or computer by voice? No longer have to get up to turn on or off a particular electrical appliance, no longer have to type for hours on a keyboard to write a report (for example). Man is constantly in search of improving quality of life, such technology has always been in him a hand envy and interest that few other technologies have been able to do [5].

The natural language processing is a vast area of research where alongside specialists from many disciplines: linguists, computer scientists, logicians, psychologists, translators... It is also an economically area holder whose

numerous applications in areas as diverse as office, assistance for the disabled, education, home automation, translation, assistance navigation, documentation, not to mention computing. The use of recognition speech thus has a very wide range of application that requires most further research. The evolution of technology has enabled the emergence of software-and large public solutions [6].

How does speech recognition work?

The mode of speech recognition is relatively easy to learn; it generally involves an essential initial learning phase where the user adapts the software to his voice dictating a number of reports aloud. These recorded and digitized (then rendered in text format) sentences are given to program Automatic Speech Recognition (ASR).

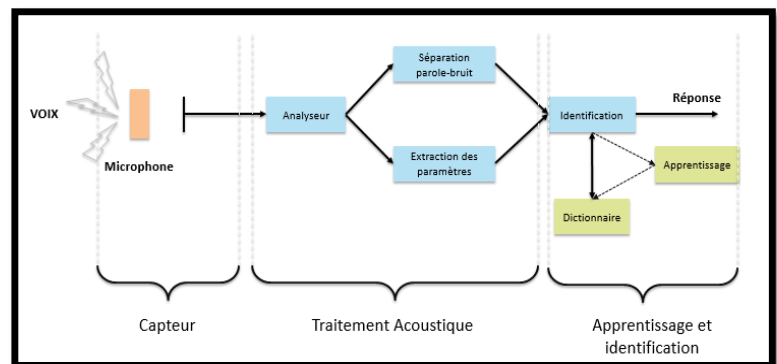


Fig 2 - Recognition Principles

Thereafter, the user reads and corrects the printed words for the changes to take effect (Fig 2.).

What constraints?

The speech recognition systems must take into account several parameters, namely:

The type of signal: noisy signal or noiseless signal (eg headset microphone with noise reduction), telephone signal (fixed or mobile) or broadband compressed signal or not ...

The type of acoustic model: model mono speaker (eg dictation), multi speaker model (speaker independent English)

The nature of records: text dictation, voice commands, man-machine dialogue, phone message, etc. ...

Language: Complexity and nuances of the language [7].

III. IMPLEMENTATION RESULT

A- General Operations

Despite numerous reports describing the operation of the speech recognition as a whole, very few actually explain its association with Android. Indeed, the literature on OpenSource solutions such as Sphinx explaining the use of

speech recognition in applications remains difficult to access. Nevertheless, the research we have done has enabled us to understand the different mechanisms of benefit surrounding this technology and to deduce the best suited to our needs architecture. Then, we opted for a board and totally independent solution not only for reasons of compliance with the constraints of the subject of study, but also an opening to a wider audience including all those interested but do not necessarily have to an internet connection quality.

Based on the above structure, our application works in two modes namely:

1- Background

When an incoming or outgoing call, the application is automatically triggered. Registrations, as well as the transcript of the call are made automatically

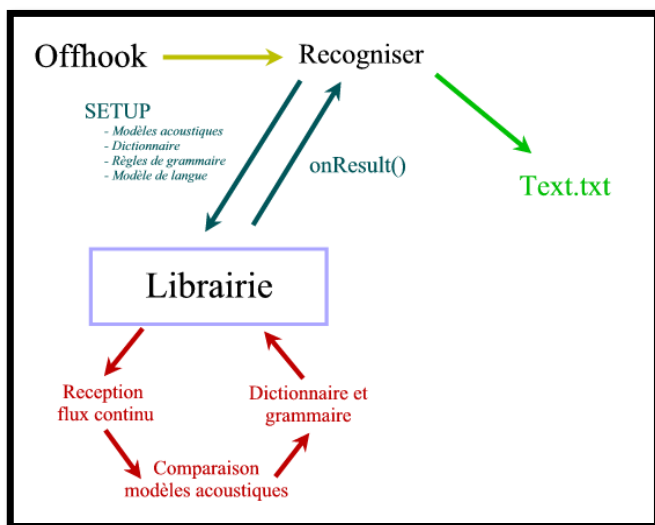


Fig 3. Recognition System

To carry out this project, we used development tools Eclipse for the code, and Bitbucket for real-time synchronization of our application. So we opted for the recognition engine Sphinx which was the heart of our production. Therefore, at the end of the call, the traces of this activity are recorded in a log file so to be returned later on user request (Fig 3.).

2- Users Interactions

We decided to make an application more ergonomic as possible. To do this, we left the possibility for each user to enable or disable each of the features of our application.

Thus, after activating the translation of the text report obtained and selected the target language of the translation, a connection is made with Microsoft Translate servers to ensure spots translation and present them to the user (Fig 5.).

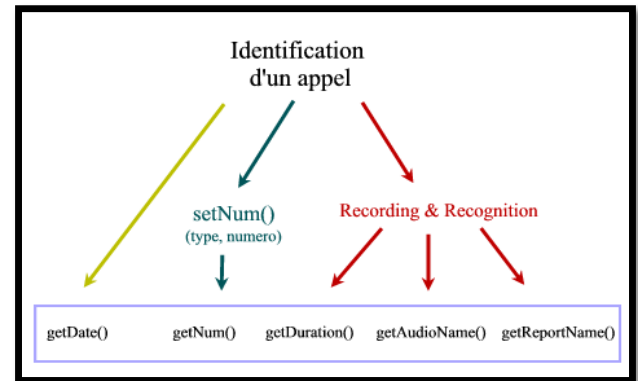


Fig 4. Log File

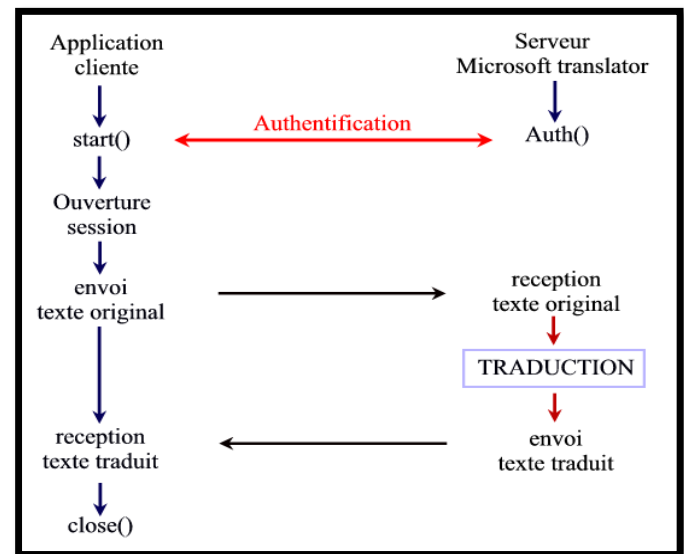


Fig 5. Microsoft Translator Service

As showing in the figure Fig 5, the report reflects obtained may therefore be subject to an option to share via email, SMS, or otherwise. Also the user can listen to the recordings will [8].

B- Interaction and communication structure

1- Technology Infrastructures

Thus, in the context of our application, several modules interact to provide all the services that we offer. Thus, we can note inside our application tool call transcription, voice recorder GSM telephone and a text translator (several languages). These modules are interconnected, and are mostly activated upon detection of an incoming or outgoing call [9]. And conversations are recorded, transcribed and stored in the phone memory. Interface summarizing all registrations allow access to transcripts and translation when the latter option is chosen.

Too late to carry out this project and in order to have a wide range of acoustic models, we must, among other things have a recording studio and tools for processing voice (Fig 6.).

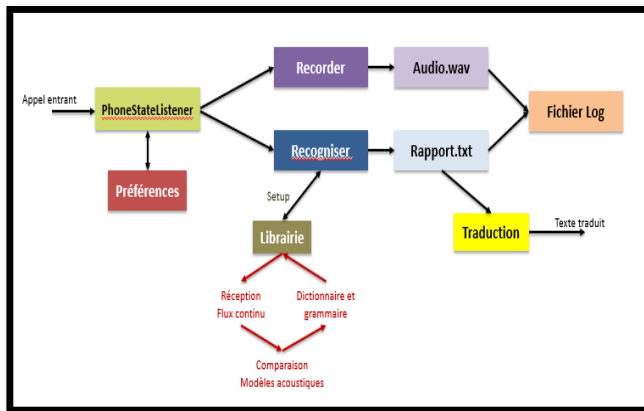


Fig 6. Implementation of the application

2- Privacy and Security issue

To reduce the risk of interdiction between mobile application and server, we opted for a fully embedded and standalone solution. Consequently, only requests to translate the written language into a target report will be sent to a remote server, the Microsoft. This secure connection requires prior authentication ensures that trade with Microsoft servers [12].

In addition we integrate the perspective of an electronic signature report to prevent it being tampered with. As we envision a security module of the application password. The user can enter a pin code to be authenticated each time you open the application.

IV. ASSESSMENT AND PROSPECTS

The use of speech recognition has a very wide range of application that requires mostly still research [13]. The evolution of technology has led to the emergence of fully software solutions and large audiences. However, the implantation of an automatic speech system in real conditions pose many problems such as noise, the state of the speaker, the recording quality, the difficulty of vocabulary, mode of speech, the system integration in a device. And develop an application for transcribing telephone allowed us to really discover the below technologies speech processing.

Furthermore, the maintenance phase of the application will allow us to incorporate corrections into account the opinions and criticisms to improve the application in its general operation. Correcting any bugs, expanding the field of languages and adding additional services, we

propose and services increasingly tailored to the needs of users. Also it would be good to consider adding an authentication when opening the application (security aspects) and a module to have a voicemail to which the caller can leave a message. This voice message left can then be transcribed and presented to the user, once it available in the form of missed notification.

V. CONCLUSION

Speech recognition is a computer technique for analyzing speech samples from a source audio and transcribe text as readable by a machine. Allow a user to apply orally to a machine for various tasks it aims to: transcription, control, translation, and is mainly based on several scientific fields, namely the theory of information, signal processing, automatic language processing, linguistics, artificial intelligence and many others. Therefore, the implementation of an automatic speech in real conditions implies many problems such as noise, the state of the speaker, the recording quality, the difficulty of vocabulary, and mode of speech the integration of the system equipment.

However, with the electronics development of and the computing great boom these recent years, several projects have emerged, aiming to achieve a system that can transcribe in real time with the minimum of error human speech. These projects will lead the various software solutions that we find on the market today.

REFERENCES :

- [1] ASR, "Voice Standards", 117 pages, ISBN-10 9241547219, December 2007.
- [2] "Information Economy Report 2009" http://unctad.org/en/docs/ier2009_en.pdf, last accessed on Feb 2012.
- [3] "The Millennium Development Goals Report 2011" http://www.un.org/millenniumgoals/11_MDG%20Report_EN.pdf, last accessed on Feb 2012.
- [4] "ICD-10 Version:2010", <http://apps.who.int/classifications/icd10/browse/2010/en#/I on the 5th of March 2012>.
- [5] "Applying ICD-10 to verbal autopsy", http://www.who.int/healthinfo/statistics/verbal_autopsy_standards3.pdf, last accessed on the 11th of Mai 2012.
- [6] The WHO Young Infants Study Group. Clinical prediction of serious bacterial infections in young infants in developing countries. *Pediatric Infect Dis J* 1999;18:S23–31.
- [7] Anker M et al. A standard VA method for investigating causes of death in infants and children. Geneva, World Health Organization, 1999 (WHO/CDS/CSR/ISR/99.4; http://whqlibdoc.who.int/hq/1999/WHO_CDS_CSR_ISR_99.4.pdf, last accessed on the 3th Jun 2012.
- [8] "UML Extension for Objectory Process for Software Engineering" version 1.1 (1 September 1997). The Object Management Group, doc. no. ad/97-08-06.
- [9] "Mobile Information Device Profile (MIDP)", <http://jcp.org/en/jsr/detail?id=118>, last accessed Jan 2012.
- [10] I. Salmre. Essential Software Engineering for Building Mobile Applications. Pearson Education, Inc. ISBN 0-321-26931-4. First printing, February 2005.

- [11] J. Knudsen. Creating Great Mobile Applications. Safari Books online. ISBN 0-321-4642-0. First printing, December 2012.
- [12] Becker SR, Diop F, Thornton JN. Infant and child mortality in two counties of Liberia: Results of a survey in 1988 and trends since 1984. International Journal of Epidemiology, 1993, 22(1):S56-63.
- [13] Anker M. The effect of misclassification error on reported cause-specific mortality fractions from verbal autopsy. International Journal of Epidemiology, 1997, 26(5):1090-1096.



Mr. James K. Tamgno, born in September 1973, I have been a Ph.D student since October 2010. The tentative title of my thesis is "Promotion of African's languages and dialects and Contribution to a Wider Accessibility of ICT, by Circumventing Language Barriers and Lessening the Impact of Illiteracy or Disability".

Previously I have gotten a M.Sc. in Engineering Science from University Cheikh Anta DIOP of Dakar, studied numerical analysis and mathematics at the University of Yaoundé, and worked as a research scientist at ESMT (Multinational High School of Telecommunication) of Dakar.
<http://www.esmt.sn/pages/personnel.php?c=J>



Kibanou Eric TIDIYE was born in Lomé, Togo, in 1993.

He receive bachelor's degree from Collège Protestant , Lomé, Togo in 2010. In 2013 after two months of internship in TOGO Telecom working in Network Operating Center, he receive "Diplome de Technicien Supérieur" at ESMT,Dakar,Senegal . He received Bachelor in Network and Services after working two months

on the call transcription system for mobile phones with Loic AKOHA in 2014 from ESMT,Dakar,Senegal. Since Febuary 2015 he is working at ART&P (Regulatory Authority of Posts and Telecommunications) as trainee on mobile application able to perform quality of service test for mobile network, Lomé, Togo.



Loic AKOHA was born in Cotonou, Benin, in 1993.

He receive bachelor's degree from Collège Catholique Père Aupiais, Cotonou, Benin in 2011. First, he got his "Diplome de Technicien Supérieur" after two years of studies in teleinformatic at ESP, Dakar,Senegal . In 2014, he received Bachelor in Information and

Communication Technologies after working two months on the call transcription system on Android mobile phones with Kibanou Eric TIDIYE from ESMT,Dakar,Senegal. He is currently pursuing his studies at UQAM in Montreal, Canada (QC), where he is registered in Master of Science in Informatic.